Fitness Effects of Changing Codon Bias of two genes – *mauA* and *mtdA* in *Methylobacterium extorquens AM1*

Nilima Walunjkar

Indian Institute of Science Education and Research, Pune

Biology

Advisor:

Dr. Deepa Agashe

National Centre for Biological Sciences

Certificate

This is to certify that this dissertation entitled “Fitness Effects of changing codon bias of two genes – *mauA* and *mtdA* in *Methylobacterium extorquens*” towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Nilima Walunjkar at the National Centre for Biological Sciences under the supervision of Dr. Deepa Agashe, Principal Investigator, during the academic year 2016-2017.

Nilima Walunjkar
BS-MS student
IISER Pune

Dr. Deepa Agashe
Principal Investigator
National Centre for Biological Sciences

Declaration

I hereby declare that the matter embodied in the report titled “Fitness effects of changing the codon bias of two genes – *mauA* and *mtdA* in *Methylobacterium extorquens*” are the results of the work carried out by me at the National Centre for Biological Sciences, under the supervision of Dr. Deepa Agashe and the same has not been submitted elsewhere for any other degree.

Nilima Walunjkar
BS-MS student
IISER Pune

Dr. Deepa Agashe
Principal Investigator
National Centre for Biological Sciences

Abstract

Organisms have a codon bias i.e. they use certain codons more frequently than others that code for the same amino acid. Studies have shown that changing the codon bias of a gene can have a negative impact on fitness due to decreased mRNA and protein levels. Mechanisms like tRNA-codon imbalance, ribosome pausing and altered mRNA secondary structure have been proposed to explain the fitness effects of altering codon usage. We are studying the fitness effects of changing the codon bias of two genes, *mauA* and *mtdA* in *Methylobacterium extorquens.* Changing the codon bias of *fae,* another gene in *M. extorquens* has a negative impact on fitness. By comparing the trends in fitness effects across genes from the same pathway, we may unearth the mechanistic basis of this effect. We generated codon usage variants of *mauA* and *mtdA* and quantified their growth rates. Preliminary analysis shows that fitness effects in *mauA* strains are similar to those of *fae.* While those seen in *mtdA* codon usage variants appear to depend on the carbon substrate available, though this may be an artefact of the FLAG tag introduced into the sequence. Also, a previous study in the lab evolved the codon altered *fae* strains and sequenced them. Here we analyse this sequence data to identify background mutations that could contribute to increasing fitness. Our preliminary data supports the hypothesis that fitness effects of synonymous mutations are context dependent.

List of Figures

1. Effect of tRNA-codon usage imbalance on protein production 8
2. Altering mRNA secondary structure interferes with initiation of translation 9
3. Presence of SD like sequences in the transcript leads to ribosomal pausing and mis-initiation 10
4. The formaldehyde oxidation pathway of M. *extorquens* 12
5. The construct to be cloned into pCM433 16
6. A schematic of the triparental conjugation protocol 18
7. Growth rates of *mauA* strains 22
8. Growth rates of *mtdA* strains 23
9. Growth rates of *fae* strains 24

List of Tables

1. Composition of Hypho 14
2. Composition of metal mix 14
3. Codon composition of *mauA* variants 15
4. Codon composition of *mtdA* variants 15
5. Primers used in this study 17
6. pCM433 plasmids used in this study 20
7. Unique SNPs in the evolved *fae* lines 26
8. Unique Indels in the evolved *fae* lines 27
9. SNPs and Indels common to all the evolved *fae* lines 28

Acknowledgements

I wish to thank my mentor Dr. Deepa Agashe, for guidance, and the opportunity to work on this project. I would like to express my gratitude to Mrudula Sane for illuminating discussions and critical comments. I am grateful to Gaurav Diwan, Kruttika Phalnikar and Dr. Laasya Samhita for advice and discussions regarding experiments. I would also like to thank all the members of the Agashe lab for feedback on the project. I am grateful to Dr. Deepak Barua for his valuable inputs on this work. I would like to acknowledge Alefiyah Habibullah's contribution to this project. Finally, I thank INSPIRE and NCBS for financial and laboratory support.

Introduction

The genetic code is degenerate which means that a single amino acid can be coded for by multiple codons. Organisms use certain codons more frequently than others to code for an amino acid. This is termed as 'codon bias' and is a genome level property of a species. Changing the codon bias of a gene is essentially equivalent to making numerous synonymous mutations (mutations that change the nucleotide sequence of the coding region, but do not change the amino acid coded for) in the gene. As they leave the protein structure and function intact, it was expected that synonymous mutations would be neutral with respect to fitness (Kimura, 1968; King and Jukes, 1969).

However experimental evidence across bacteria has shown that this is not true. Agashe et al. (2013) changed the codon bias of an essential gene in *Methylobacterium extorquens* AM1 and observed that it reduced fitness. Hauber et al. (2016) demonstrated that single synonymous mutations in a ribosomal protein of *Escherichia coli* reduced its fitness. Lind et al. (2010) made single synonymous mutations in two ribosomal proteins, *rpsT* and *rplA* of *Salmonella typhirium* and found them to have weakly deleterious effects. Bailey et al. (2014) found two beneficial synonymous mutations arising in their experimentally evolved lines of *Pseudomonas fluorescens*. The fitness advantage that these synonymous mutations conferred was comparable to other non-synonymous mutations that arose.

There exist multiple mechanistic explanations of how synonymous mutations can impact fitness. These are summarized below.

1. tRNA-codon usage imbalance:

   This theory posits that the codon usage and the tRNA pool of a genome have co-evolved to optimize the efficiency of translation (Bulmer, 1988; Ikemura, 1985). The codon preference model suggests that there is weak selection for the usage of the most optimal codon at every site, but mutation and genetic drift maintain the non-optimal codons in the genome (Bulmer, 1991). This balance between selection, mutation and drift is what shapes and maintains the codon bias of a genome.

It can so happen that a synonymous mutation introduces a rare codon into the coding sequence of a gene. During translation of its mRNA, the ribosome pauses at this rare codon until it encounters the corresponding cognate aminoacyl-tRNA (Curran and Yarus, 1989; Plotkin and Kudla, 2011). Ribosome sequestration reduces the efficiency of translation and leads to reduced protein levels and lowered fitness (Plotkin and Kudla, 2011) (Figure 1). Studies have also shown that highly expressed genes tend to be codon optimized (Gouy and Gautier, 1982). Hence the expectation is that changing the codon bias of highly expressed genes will have larger effects on fitness.

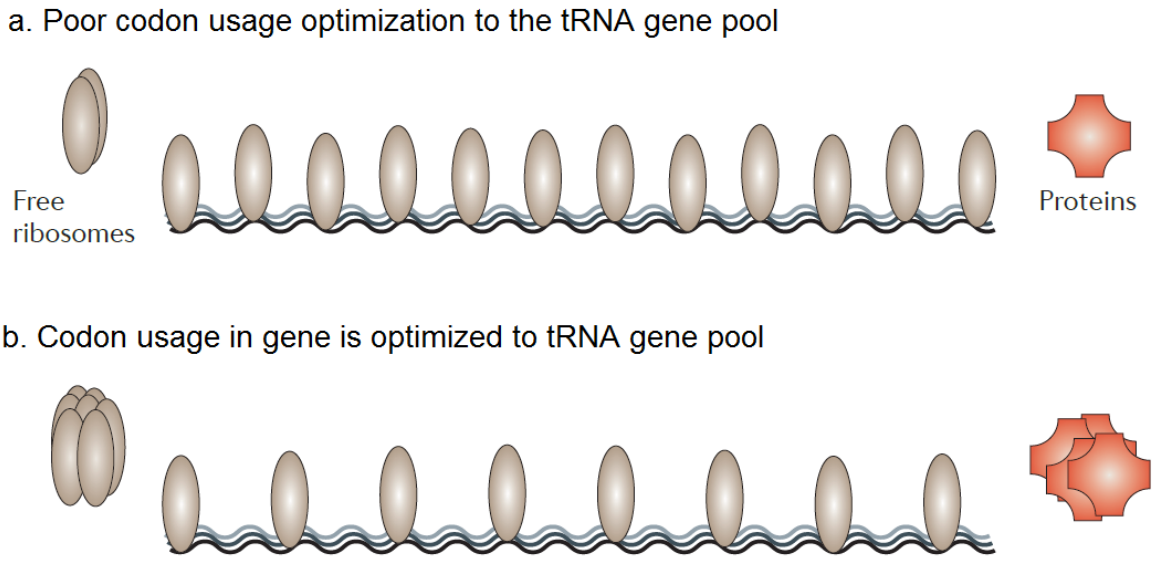

Figure 1: Effect of tRNA-codon usage imbalance on protein production. (a) If the codon usage of the gene is not optimized to the available tRNA gene pool, (for example, an exogenous gene after horizontal gene transfer) the ribosomes will be sequestered on the mRNA and this lowers protein production. (b) On the other hand when codon usage is optimized to the tRNA gene pool, there is efficient translation and high levels of protein production. Adapted from Plotkin and Kudla (2011).

2. Altered mRNA secondary structure:

Synonymous mutations alter the nucleotide sequence and hence the secondary structure of the transcribed mRNA (Kelsic et al., 2016; Lind et al., 2010). An altered mRNA secondary structure can interfere with the initiation of translation by preventing the binding of the ribosome to the Shine-Dalgarno

sequence (Li et al., 2012) (Figure 2). The Shine-Dalgarno (SD) sequence is the motif on the transcript bound by the ribosome to initiate translation. Changing mRNA secondary structure can also lead to sequestration of ribosomes on the transcript (Kudla et al., 2009). Another possibility is that the altered secondary structure is unstable and this leads to increased degradation of the transcript, ultimately lowering protein levels (Presnyak et al., 2015).

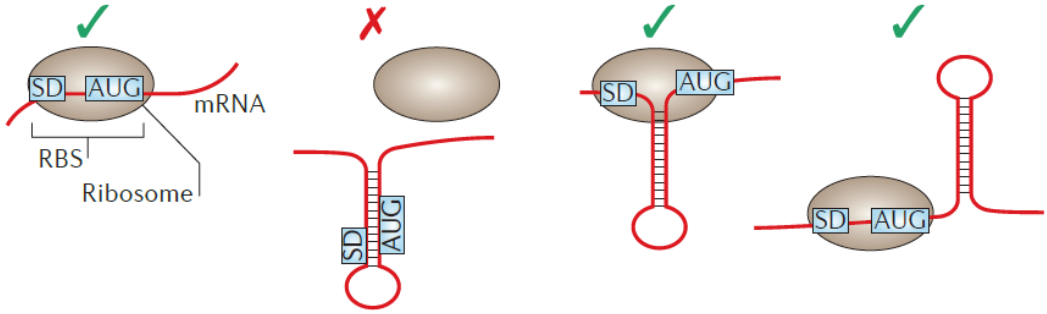

Figure 2: Altering mRNA secondary structure can interfere with initiation of translation by preventing the binding of the ribosome to the Shine-Dalgarno sequence, thus leading to lower protein levels. Figure from Plotkin and Kudla (2011).

3. Introduction of Shine-Dalgarno (SD) like sequences:
   Normally, strong SD sequences occur only upstream of genes. Introducing synonymous mutations in a gene can lead to the formation of Shine-Dalgarno (SD) like sequences which exhibit an increased affinity for the anti-SD sequence in the ribosome. This can in turn lead to mis-initiation or pausing of ribosomes at SD like sequences within genes (Li et al., 2012) (Figure 3). Both mis-initiation and ribosomal pausing will lower protein levels and thus impact fitness.

Recent studies have attempted to unearth the mechanistic basis of how synonymous mutations affect fitness. Agashe et al., (2013) changed the codon bias of the *fae* gene in *Methylobacterium extorquens* AM1. *M. extorquens* can utilise single carbon resources like methylamine and methanol as well as multi-carbon substrates like succinate to derive energy. *Fae* (Formaldehyde activating enzyme) is an enzyme in the formaldehyde oxidation pathway of *M. extorquens* that catalyses the condensation of formaldehyde with H_4PMT (tetrahydromethanopterin) (Marx et al., 2003). Altering the codon bias of *fae* lowered its growth rate when the sole recarbon source available was methanol. This was attributed to lower levels of the enzyme and the *fae* transcript.

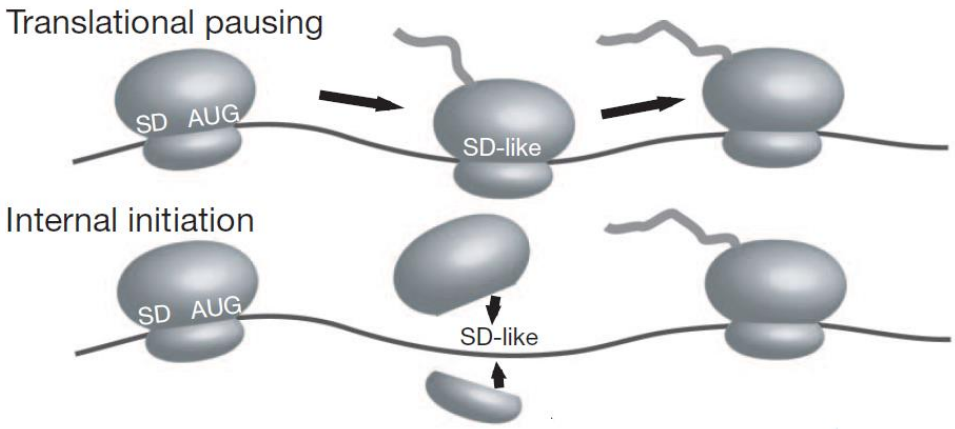

Figure 3: Presence of SD like sequences in the transcript can lead to ribosomal pausing and thus sequestration of ribosomes. Alternately it can also lead to mis-initiation from the SD like sequence. Figure from Li et al., (2012)

They generated six strains to cover a broad range of codon bias and positional effects of the codons. These included ones in which every codon of *fae* was a frequent codon (AF: All Frequent) and every codon of *fae* was a rare codon (AR: All Rare). The AR allele is not optimized to the tRNA gene pool of *M. extorquens* AM1. If tRNA-codon usage imbalance played a role, AR was expected to have low fitness and the fitness of AF would be higher than wild-type or comparable to it. Contrary to this expectation, they observed that the AF strain had very low fitness and the AR strain had a higher fitness than AF but lower than wild-type. The protein levels of *fae* in all their strains did not correlate with the mRNA folding energy of the codon altered variants. They found a weak negative correlation between *fae* protein levels and SD like sequences that occurred within the altered *fae* alleles. Thus their results did not provide unequivocal support for any of the proposed mechanistic explanations. Hence they concluded that effect of synonymous mutations is context dependent i.e. it depends on the gene and the position of the mutation within the gene.

Hauber et al. (2016) designed mutants of *Escherichia coli* to test the positional effects of synonymous mutations. They changed two lysine and two leucine codons to the corresponding rare codon. These were located sufficiently away from the 5' end of the transcript to eliminate effects on translation initiation. They observed that the fitness effects of changing codons was position dependent. They also constructed a strain that had five synonymous mutations and this showed a large reduction in fitness. So

they concluded that negative synergistic epistasis between different synonymous mutations lead to reduced fitness in this mutant. The tRNA-codon usage imbalance hypothesis predicts that the effect of changing a frequent codon to a rare codon should be independent of its position within the gene. The results of this study provided evidence contrary to this prediction and the authors concluded that the effects of synonymous mutations may be position dependent.

Lind et al. (2010) made synonymous mutations in two non-essential ribosomal proteins in *Salmonella typhirium* and measured the distribution of their fitness effects. They observed that most of these mutations were weakly deleterious and some were neutral. In a follow up study, they restored the fitness of three out of seven mutants by making a second mutation that alleviated the effect of the first on the mRNA secondary structure (Lind and Andersson, 2013). They concluded that mRNA secondary structure was important but it might not be the only mechanism by which synonymous mutations affect fitness.

The above studies find similar effects of synonymous mutations across species of bacteria. All three find that tRNA-codon usage imbalance probably does not play a major role in affecting fitness when the codon composition of only one gene is changed. They implicate mRNA secondary structure changes as being a plausible mechanism and conclude that the effects of synonymous mutations could be context dependent. If we collect data on other genes in the same pathway and compare across genes, we may unearth any broad patterns that exist and this may lead us to a mechanistic understanding of the processes that result in synonymous mutations affecting fitness. If no broad patterns emerge, the claim that the effects of synonymous mutations is context dependent would be true. The aim of this thesis is to alter the codon bias of two genes of *Methylobacterium extorquens* AM1 that are in the same pathway as *fae* and quantify their fitness, mRNA and protein levels. By comparing trends in these quantities with those observed in the *fae* strains we hope to find some indication of the mechanism underlying the fitness effects of synonymous mutations.

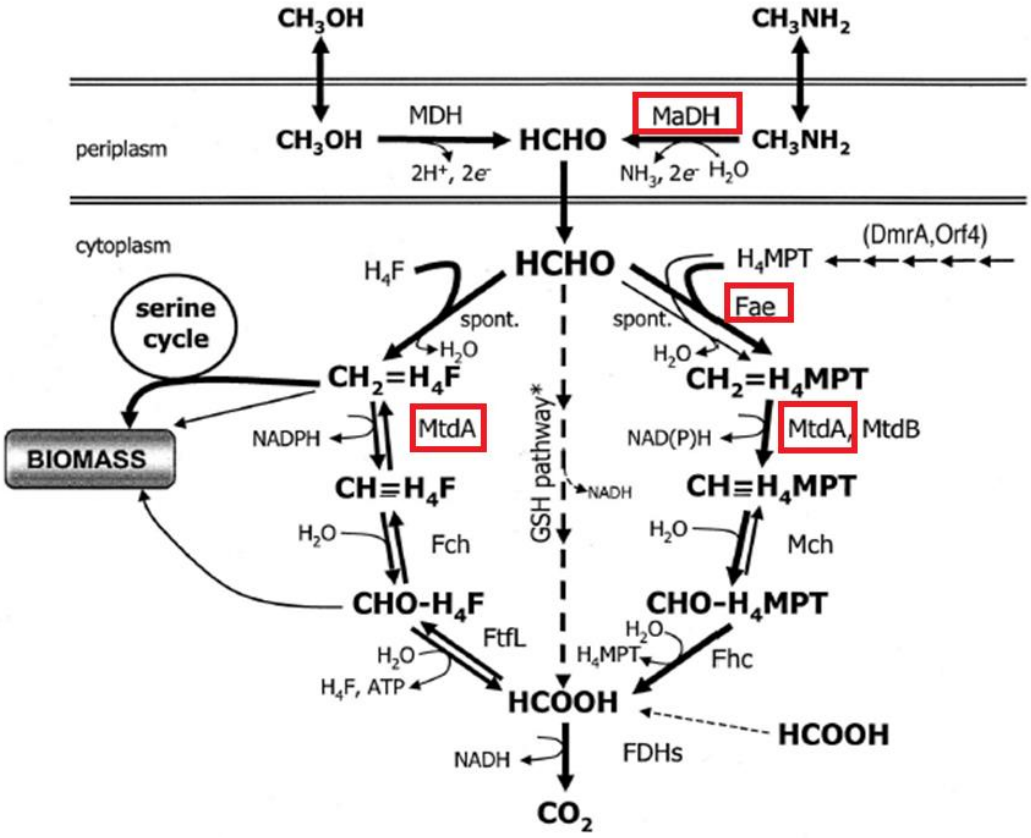

Figure 4: The formaldehyde oxidation pathway used by *M. extorquens* to assimilate single carbon substrates like methylamine and methanol. *Fae*, *mtdA* and *mauA* (MaDH) are essential enzymes in this pathway. They are highlighted in red in above. Figure adapted from (Marx et al., 2003).

Methylobacterium extorquens can use single carbon substrates like methanol and methylamine to derive energy in addition to using multi-carbon substrates like succinate. Its formaldehyde oxidation pathway (Figure 4) can assimilate methylamine and methanol to generate energy. *Fae, mauA* and *mtdA* are a part of this pathway. *MauA* codes for an enzyme that converts methylamine to formaldehyde (Marx et al., 2003). *MtdA* acts further downstream and is an NADP dependent dehydrogenase that oxidizes methylene H_4PMT to methenyl-H_4PMT (Marx et al., 2003). *Fae* and *mtdA* are highly expressed constitutively even in the absence of single carbon substrates (Bosch et al., 2008). On the other hand, the expression of *mauA* is induced only in the presence of methylamine (Bosch et al., 2008). As all three genes are highly expressed, they are expected to be codon optimized by selection and the impact of altering their codon bias is expected to show up readily. Consistent with this expectation, *mauA* and

mtdA are composed of a higher percentage of frequent codons compared to rare codons (Table 3 and 4).

I. Generating M. extorquens AM1 *strains with codon usage altered mauA and mtdA alleles*

The first question we address is whether the effects of synonymous mutations are context dependent. We generate mutants of *M. extorquens* AM1 in which the wild-type gene (*mauA* or *mtdA*) is replaced by an allele with altered codon usage. We then measure growth rates of these strains and compare across genes to discern any patterns that emerge. If a growth defect is seen, we plan to quantify mRNA levels and protein levels of *mauA* and *mtdA*.

II. Generating strains with mauA alleles on the inducible plasmid pLC291

Replacing the wild-type allele in the *Methylobacterium extorquens* AM1 genome with a codon bias altered allele is a time consuming procedure with very low efficiency (See Results). An alternative approach is to transform a plasmid carrying the codon-altered allele into *M. extorquens* AM1 in which the corresponding gene has been knocked out. Chubiz et al., (2013) developed an inducible expression vector in which the gene can be cloned under an inducible promoter. The aim is to clone the variant alleles into this plasmid, transform it into *M. extorquens* AM1 *ΔmauA* and assay fitness.

III. Analysing background mutations in the evolved strains with altered codon usage of fae

The strains with altered codon bias in *fae* showed lower growth rates than wild-type (Agashe et al., 2013). These strains were evolved till the fitness recovered close to wild-type levels (Agashe et al., 2016). When the *fae* gene in the evolved lines was sequenced, it was seen that compensatory point mutations arising near the 5' end of the gene restored fitness to almost wild-type levels (Agashe et al., 2016). When these mutations were transferred to the corresponding ancestral strains, it was observed that they only contributed partly to the recovery in fitness. It is likely that some background mutations (mutations occurring outside the *fae* gene) may contribute to the increase in fitness of the evolved lines. Previously in the lab, whole genome sequencing of isolates of the evolved *fae* lines had been done. The aim is to analyse this sequence data, identify background mutations, confirm their existence and characterize them.

Methods

Media and Growth conditions

The *Methylobacterium* growth medium used in this study is Hypho. The composition of the medium is given in Table 1. Hypho was supplemented either with 5% w/v sodium succinate or 15 µM methylamine as required. Solid media was prepared with 2% w/v agar and Hypho. All cultures and plates (including those for growth rate measurements) were incubated at 30^{o}C with 180 rpm shaking when required.

I. Generating M. extorquens AM1 strains with codon usage altered mauA and mtdA alleles

Six variants of the gene with altered codon usage were chemically synthesized by Integrated DNA technologies and inserted into the plasmid pSMARTHcKan. These variants and their codon usage for both the genes is summarized in Table 3 and 4.

Components		Final concentration in 1L media	To constitute 1L of media
P solution	$K_2HPO_4.3H_2O$	14.5 mM	100 mL of 10X P solution
	$NaH_2PO_4.H_2O$	18.8 mM	
S solution	$(NH_4)_2SO_4$	3.8 mM	100 mL of 10X S solution
	$MgSO_4$	0.8 mM	
Metal Mix	See Table 2		1mL

Table 1: Composition of Hypho (as given in Delaney et al., 2013)

Components	Volume/amount
Distilled H_2O	500 mL
EDTA	5 g (or 6.37 g of EDTA.$2H_2O$)
$ZnSO_4.7H_2O$	2.2 g
$CaCl_2.2H_2O$	0.733 g
$MnCl_2.4H_2O$	0.506 g
$FeSO_4.7H_2O$	0.499 g
$(NH_4)MO_7O_{24}.4H_2O$	0.110 g
$CuSO_4.5H_2O$	0.157 g
$CoCl_2.6H_2O$	0.161 g

Table 2: Metal Mix composition. pH is adjusted to 5.0 after addition of each component (as given in Delaney et al., 2013).

mauA	Strains	% of rare codons	% of frequent codons	Number of Synonymous mutations
WT	Wild-type	8.6	59	0
AF	All Frequent	0	95.6	69
AR	All Rare	89.2	0	163
RN	Randomly picked sites rare	44.6	47.3	118
VA	Variable sites rare	45.7	46.8	112
CO	Conserved sites rare	45.7	46.8	124
AC	Active sites rare	9.7	86	81

Table 3: Codon usage variants of *mauA*

mtdA	Strains	% of rare codons	% of frequent codons	Number of Synonymous mutations
WT	Wild-type	5.9	69.8	0
AF	All Frequent	0	97.6	82
AR	All Rare	94.1	0	265
RN	Randomly picked sites rare	46.5	49.7	172
VA	Variable sites rare	47.2	49	172
CO	Conserved sites rare	47.2	48.6	171
AC	Active sites rare	7.6	89.9	96

Table 4: Codon usage variants of *mtdA*

The first variant is the wild-type (WT). In the second variant (AF), all codons were replaced by the most frequent codon for each amino acid. Similarly, the third variant (AR) has rare codons at every position. The fourth variant (RN) was constructed with rare codons at 50% randomly picked sites and frequent codons at the remaining 50% sites. The fifth variant (VA) is composed of 50% rare codons in the variable regions of the gene while the sixth variant (CO) has 50% rare codons in the conserved sites of the gene. In the seventh variant (AC) the codons at the active sites of the enzyme have been replaced by rare codons. A short FLAG tag has been inserted at the N terminal or the C terminal of the protein to help with protein quantification at later stages. The neutrality of this tag will be determined before fitness assays are carried out.

The wild-type allele with the FLAG inserted at either the N terminal or C terminal of the protein is used to check the neutrality of the FLAG and serves as the control for this experiment. AF and AR can be used to test the tRNA-codon usage imbalance

hypothesis. If this explanation holds, the fitness of AR must be very low and AF must do either as well as or slightly better than WT. RN, VA and CO have similar codon composition i.e. they are composed of 50% frequent codons and 50% rare codons. The difference between the three strains is the position at which rare codons are located. Their fitness data can be used to determine if the effect of changing codons depends on the position of rare codons – in the conserved regions, the variable regions or in randomly picked positions. Fitness assays of AC will reveal any positional effects of having rare codons at the active sites of the enzyme.

We replaced the wild-type allele in the genome of *M. extorquens* AM1 with one of these seven variants using the method described by Marx (2008) (Figure 6). We made a construct in the plasmid pCM433 that has the variant allele flanked by the genomic upstream and downstream region (~600 bp) of the gene (Figure 5). The plasmid pCM433 has a tetracycline resistance gene, a *sacB* gene which confers sucrose sensitivity in addition to having a Multiple Cloning Site (MCS). We first amplified 600bp region just upstream and downstream of *mauA* with primers that introduced restriction sites into the amplicon. We inserted the upstream region into pCM433 after digesting the plasmid with *XbaI* and *PstI* and the amplified upstream region with *NheI* and *PstI.* The downstream region was inserted into pCM433 with the upstream region after digestion of both with *ApaI* and *AgeI*. The *mauA* alleles were ligated into this construct after digestion of both with *PstI* and *MluI*. We removed the N terminal FLAG from this construct by digestion with *NdeI* and the C terminal FLAG by digestion with *NotI* when required. The primers used for this are summarized in Table 5. The pCM433 plasmid with the upstream and downstream regions corresponding to *mtdA* already existed in the lab. We ligated the *mtdA* variant alleles into this plasmid after digestion of both with *PstI* and *MluI*.

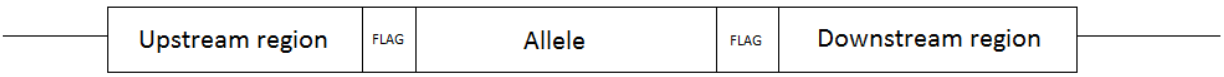

Figure 5: The construct to be cloned into pCM433. It consists of the allele with the altered codon usage flanked by approximately 600 bp of upstream and downstream region from the genome. FLAG is inserted at the N terminal and the C terminal of the protein.

Sl. No.	Primer name	Primer sequence (5' to 3')	Purpose
1	DA 64F	TCCATCCAGCAGTTCATGACG	Confirming allele replacements of *mauA* by Sanger sequencing
2	DA 64F	TGATGCTTGGCTATGAACATGCG	
3	DA 58F	ACGAAAAGACTCTCGAGGGC	Confirming allele replacements of *mtdA* by Sanger sequencing
4	DA 58R	CCTCCCCTGTCAGTTCTTGT	
5	DA23F	ACTGCAGCGCATATGCTCGGAAAATCCCAATTCGACGATC	Cloning primers for amplifying *mauA* WT
6	DA23R	GACGCGTGCGGCCGCGCTCGCCTTGCCGACGATC	
7	DA24F	AGCTAGCCCGACGCTGCTGATGTTC	Cloning primers for amplifying *mauA* upstream region
8	DA24R	TGGGCCCTGCAGCCTTGTCATCGTCATCCTTGTAGTCCAT ATGTTCCTCGTCAGTTTTATCGGGTGCTG	
9	DA25F	AGGGCCCACGCGTGGACTACAAGGACGATGACGATAAGG CGGCCGCGTGATCAATCTCCTTTGCGGCGTTC	Cloning primers for amplifying *mauA* downstream region
10	DA25R	CACCGGTCGGAAGTACTTACTCGACGACGACCTTG	
11	DA26F	ACTGCAGCGCATATGTCCAAGAAGCTGCTCTTCCAGTTC	Cloning primers for amplifying *mtdA* WT
12	DA26R	GACGCGTGCGGCCGCGGCCATTTCCTTGGCCAGCTTG	
13	SL1	CAGTCCAGTTACGCTGGAGTC	Amplifies synonymous variant gene from pSMARTHcKan vector
14	SR2	GGTCAGGTATGATTTAAATGGTCAGT	
15	DA 349F	ATATGGTACCAGGGAGAGACCCCATATGCTGGGCAAGTC GCAGTT	To clone *mauA* AF and AC into pLC291
16	DA 349R	ATATGAATTCTCACGCGGCCGCCTTATCGTCATCGTCCTT GTAGTCCACGCGTGCGGCCGCCGA	To clone *mauA* AF, RN, CO and AC into pLC291
17	DA 350F	ATATGGTACCAGGGAGAGACCCCATATGTTGGGAAAAAG TCAATTTGATGATTT	To clone *mauA* AR into pLC291
18	DA 350R	ATATGAATTCTCACGCGGCCGCCTTATCGTCATCGTCCTT GTAGTCCACGCGTGCGGCCG	To clone *mauA* AR, VA and WT into pLC291
19	DA 351F	ATATGGTACCAGGGAGAGACCCCATATGCTGGGAAAATC GCAGTT	To clone *mauA* RN into pLC291
20	DA 352F	ATATGGTACCAGGGAGAGACCCCATATGCTGGGAAAGAG TCAATTCG	To clone *mauA* AR, RN, CO and AC into pLC291
21	DA 353F	ATATGGTACCAGGGAGAGACCCCATATGTTGGGCAAATC GCAGTTTG	To clone *mauA* AR, RN, CO and AC into pLC291
22	DA 355F	ATATGGTACCAGGGAGAGACCCCATATGCTCGGAAAATC CCAATTCG	To clone *mauA* AR, RN, CO and AC into pLC291

Table 5: Primers used in this study

In the second step, we set up triparental conjugations between *M. extorquens* AM1 (the recipient strain), an *E. coli* strain harbouring the plasmid pCM433 which carries the construct with the variant allele and an *E. coli* strain carrying the plasmid pRK2073 (which has conjugal transfer genes). This will result in the transport of the plasmid pCM433 into *M. extorquens* AM1 and we select for these transformants by plating the

culture on a Hypho-Tetracycline plate. Since the upstream and downstream region in the construct is homologous to the corresponding regions in the genome of *M. extorquens,* the plasmid gets inserted into the genome by homologous recombination. At some point of time, homologous recombination occurs again and the plasmid is excised out. Now two types of cells will be generated – those that retain the wild-type allele and those in which the wild-type allele has been replaced by the mutant allele. We picked colonies from the tetracycline plate and grew them without antibiotic selection for 36 hours to ensure that the cells get rid of the plasmid. We then plated this culture on a sucrose plate to select against cells that still harbour the integrated or the excised plasmid. We then picked colonies from the sucrose plate and replica pated on a Hypho plate and a Hypho plate containing tetracycline. The colonies that grow on only Hypho but not Hypho plus tetracycline are screened by Sanger sequencing to confirm that the wild-type allele has indeed been replaced by the mutant allele.

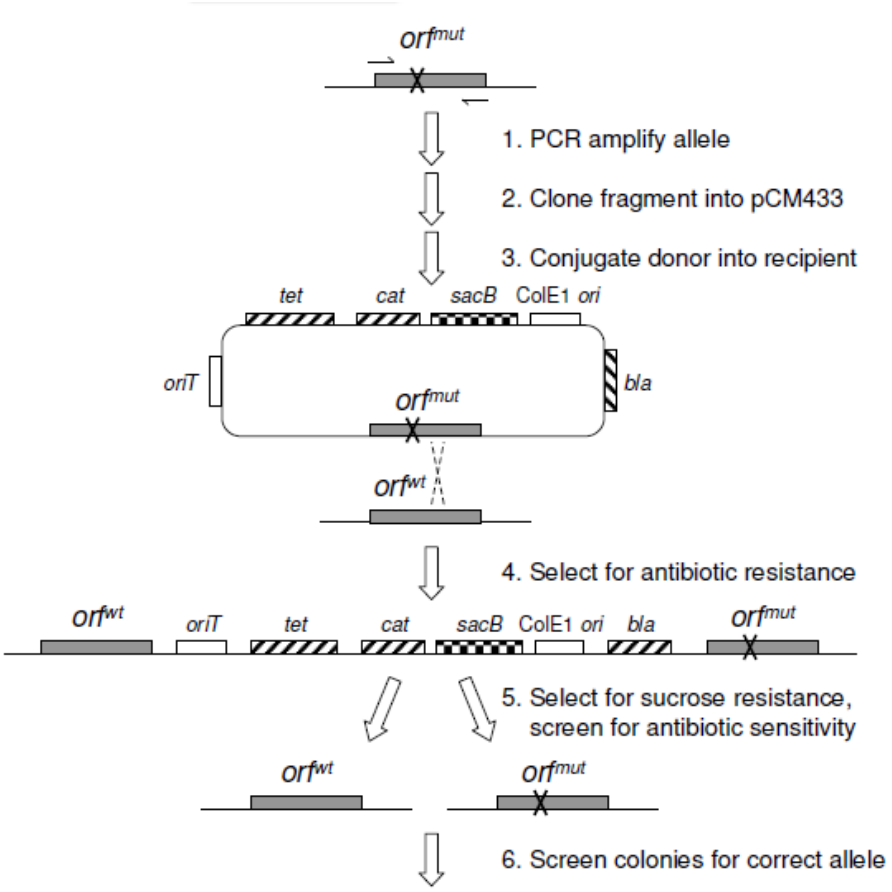

Figure 6: A schematic of the triparental conjugation protocol followed by multiple selection steps to generate *M. extorquens* strains in which the wild-type allele has been replaced by the altered allele. Figure from (Marx, 2008)

We measured the growth rate of the positive clones in the Tecan, which is an instrument that measures the OD_{600} of the culture at regular time intervals. 48 well plates were inoculated with 5 replicates of each strain in a 1:64 dilution. We fit the growth curves using Curve Fitter and calculated the growth rates and doubling times of each of the strains. The mean and standard error were plotted and the t test was used to check if the growth rates were significantly different from WT.

II. Generating strains with mauA alleles on the inducible plasmid pLC291

We cloned the *mauA* alleles with altered codon bias into the inducible plasmid pLC291 (Chubiz et al., 2013) with a ribosome binding site under the promoter $P_{R/tetO}$ that is induced by anhydrotetracycline. Triparental conjugations were then set up between *M. extorquens* AM1 *ΔmauA* (a *mauA* knockout and the recipient strain), an *E. coli* harbouring the pLC291 plasmid with the variant *mauA* allele and an *E. coli* carrying the plasmid pRK2073 (this plasmid has the conjugal transfer genes). This facilitates transfer of the plasmid pLC291 into *M. extorquens* AM1 *ΔmauA.* We will make stocks of the positive transformants that formed colonies on kanamycin plates and measured their growth rates upon induction by 100ng/ml of anhydrotetracyline. As *mauA* is essential for growth on methylamine, the growth defect of the knockout is rescued by the allele on the plasmid upon induction, although the variant alleles may do this to a different level than the wild-type allele.

III. Analysing background mutations in evolved strains with altered codon usage of fae

The whole genome sequence data of the evolved *fae* lines (already available in the lab) was analysed using VarScan (Koboldt et al., 2012, using a pipeline written by Mrudula Sane and Aalap Mogre). We aligned the sequenced genome to the *M. extorquens* AM1 reference genome (Accession no.: CP001510) from NCBI and called SNPs and indels when the following criteria were satisfied:

1. The mutation was represented on the plus strand as well as the minus strand.
2. The frequency cut-off applied was 20% i.e. if 20% of the reads showed the mutation, then it was identified as a SNP or an indel.

We compiled the indels and SNPs called for the various evolved lines and designed primers to confirm the presence of these mutations. The length of the fragment of genomic DNA amplified by these primers was 500 bp. These fragments were amplified by PCR and sanger sequenced to confirm the presence of these mutations.

Results

I. Generating M. extorquens AM1 strains with codon usage altered mauA and mtdA alleles

A. *Generating pCM433 constructs of mauA and mtdA*

We generated a construct in pCM433 with the upstream and downstream region of *mauA.* We PCR amplified the codon altered alleles from the pSMARTHcKan vector and inserted them between the upstream and the downstream region in pCM433. We transformed these into DH5α and selected for transformants by using tetracycline resistance. Plasmids with FLAG at both C and N terminal and only C terminal of the protein were generated. All the *mauA* plasmids cloned as a part of this study are summarized in the Table 6.

Strain	WT	AF	AR	RN	VA	CO	AC
mauA N and C terminal FLAG							
mauA C terminal FLAG only							
mtdA N and C terminal FLAG							
mtdA C terminal FLAG only							

pCM433 construct generated as a part of this thesis

pCM433 construct generated and moved to *M. extorquens* by Alefiyah

Construct has been moved to M. extorquens

WT	**Wild-type**
AF	**All Frequent**
AR	**All Rare**
RN	**Randomly picked sites rare**
VA	**Variable sites rare**
CO	**Conserved sites rare**
AC	**Active sites rare**

Table 6: pCM433 plasmids used in this study are tabulated above. The ones marked in blue were generated as a part of the thesis work. The plasmids marked in orange were generated by Alefiyah, a JRF in the lab. The hatched boxes represent constructs that have successfully been moved into *M. extorquens* AM1.

A plasmid with the upstream and downstream region of the *mtdA* gene was already available in the lab. The codon altered alleles of *mtdA* were inserted into this construct. Alefiyah, a JRF in the lab, had previously cloned some of these constructs into pCM433 (*mtdA* AF, VA, CO and AC with FLAG at N and C terminal and AC with FLAG at C terminal). The other *mtdA* plasmids generated in this study as summarized in Table 6.

B. *Replacing wild-type allele of M. extorquens AM1 with codon usage altered allele*

We used the triparental conjugation protocol described above (Marx, 2008) to generate *M. extorquens* AM1 strains in which the wild-type allele (*mauA* or *mtdA*) was replaced by one of the variant alleles listed in table 3 or 4. *M. extorquens* is a slow growing bacterium which takes 36 hours to reach the stationary phase in liquid culture. On a Hypho plate supplemented with succinate, it can take a minimum of four days to form visible colonies. In the presence of antibiotics, (for example, step 2 of the triparental conjugation protocol requires colonies to form in the presence of two antibiotics) *M. extorquens* can take up to eight days to form visible colonies. As the protocol has multiple antibiotic selection steps, it takes a minimum of 28 days to complete it. Additionally, the triparental conjugation protocol has a highly variable false positive rate that appears to depend on the specific construct that undergoes homologous recombination (Marx, 2008). Till date, we have set up 14 rounds of triparental conjugations. Of these, two rounds have been successful and we have generated three *mauA* strains with C terminal FLAG: WT, AR and RN and one *mtdA* strain with both C and N terminal FLAG: RN. The four *mtdA* strains generated by Alefiyah Habibullah, a JRF in the lab are: AF, VA, CO, AC with FLAGs at both C terminal and N terminal.

C. *Testing the neutrality of the C terminal FLAG (mauA) in succinate and methylamine*

Since the FLAG is not an endogenous sequence, it can disrupt translation and thus impact fitness. We tested the neutrality of the FLAG by comparing the growth rates of WT with FLAG at the C terminal of *mauA* with wild-type *M. extorquens* AM1 (Figure 7A). The growth rate of WT with FLAG was not significantly different from that of WT in both carbon substrates, succinate and

methylamine. This indicated that the FLAG is neutral and we proceeded to generate other codon altered *mauA* strains with FLAG at the C terminal. The growth rate of the *mauA* knockout, Δ*mauA,* (available in the lab) in succinate is not significantly different from WT. However, *mauA* is essential for survival of the bacterium when the sole carbon source available is methylamine (Figure

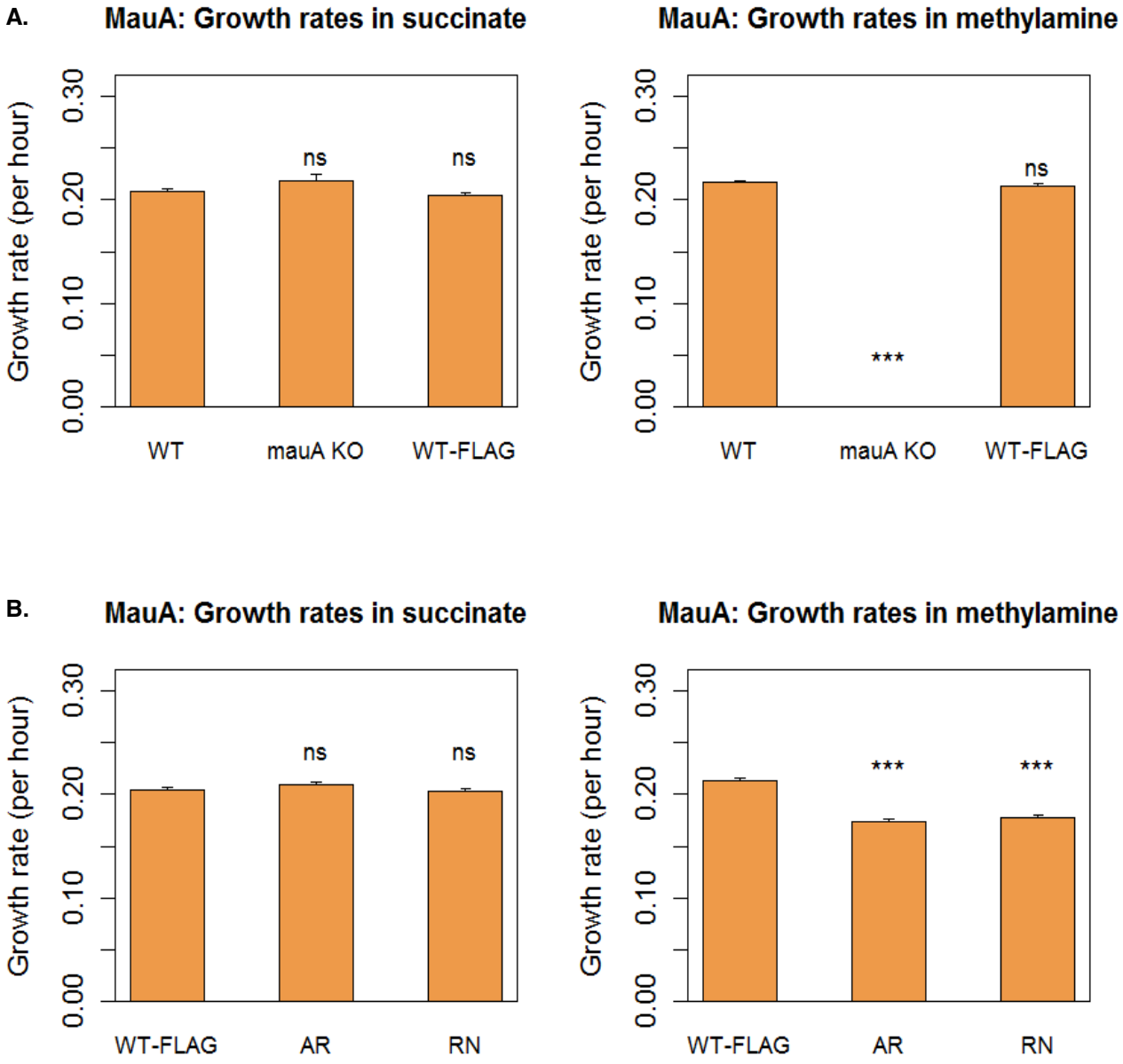

Figure 7: (A) Testing neutrality of FLAG in *mauA*: The wild-type *mauA* allele with a FLAG at the C terminal does not differ in fitness from the wild-type in both succinate and methylamine. (B) AR and RN strains of *mauA* show a growth defect in methylamine but not in succinate. Pairwise comparisons with WT (A) and WT-FLAG (B) using student's 2-tailed t test *** p< 0.001, **p< 0.01, *p < 0.05, ns – not significant. n = 5 for all strains and error bars represent standard error of mean.

7A). Thus, *mauA* is under selection when methylamine is the only carbon source available and growth defects observed in our codon altered variants can be attributed to *mauA*.

D. *Growth rates of mauA alleles with C terminal FLAG: WT, AR and RN in succinate and methylamine*

We also measured the growth rates of the AR and RN variants of *mauA.* Their growth rates in succinate are not significantly different from that of WT-FLAG. But they do show a growth defect in methylamine (Figure 7B).

E. *Growth rates of mtdA alleles with both N and C terminal FLAG: AF, RN, VA, CO and AC in succinate and methylamine*

MtdA plays a role in generating certain compounds required for biosynthesis even when *M. extorquens* is growing on succinate (Marx and Lidstrom, 2004). Thus, the *mtdA* knockout (Δ*mtdA*) can only grow on succinate supplemented with methylamine (Marx and Lidstrom, 2004). When we alter the codon usage of *mtdA,* we expect it will affect growth in both succinate and methylamine. AF,

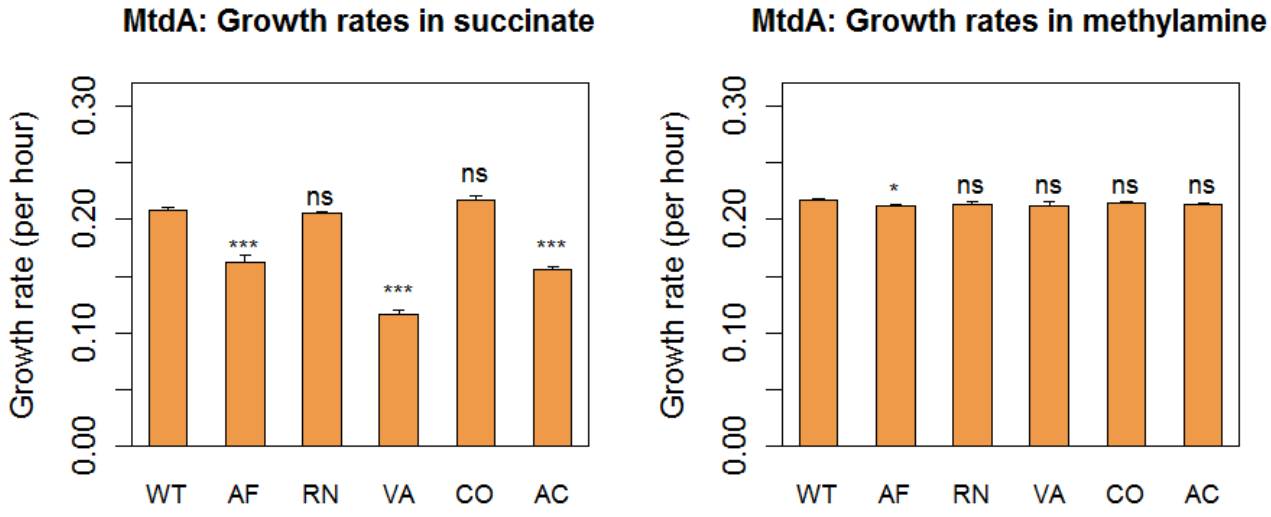

Figure 8: Growth rates of codon altered *mtdA* strains in succinate and methylamine: AF shows a growth defect in both substrates though it grows better in methylamine than succinate. VA and AC grow slower in succinate but have WT like growth rates in methylamine. Growth rates of RN and CO do not differ from WT in both succinate and methylamine. Pairwise comparisons were done with WT using student's 2-tailed t test *** p< 0.001, **p< 0.01, *p < 0.05, ns – not significant. n = 5 for all strains and error bars represent standard error of mean.

VA and AC show a growth defect in succinate compared to WT (Figure 8). While the AF strain does show a marginal growth defect in methylamine, VA and AC grow just as well as WT. On the other hand, growth rates of RN and CO are not significantly different from WT in both substrates (Figure 8).

F. Growth rates of fae alleles with C terminal FLAG: WT, AF, AR, RN, VA, CO and AC in succinate and methylamine

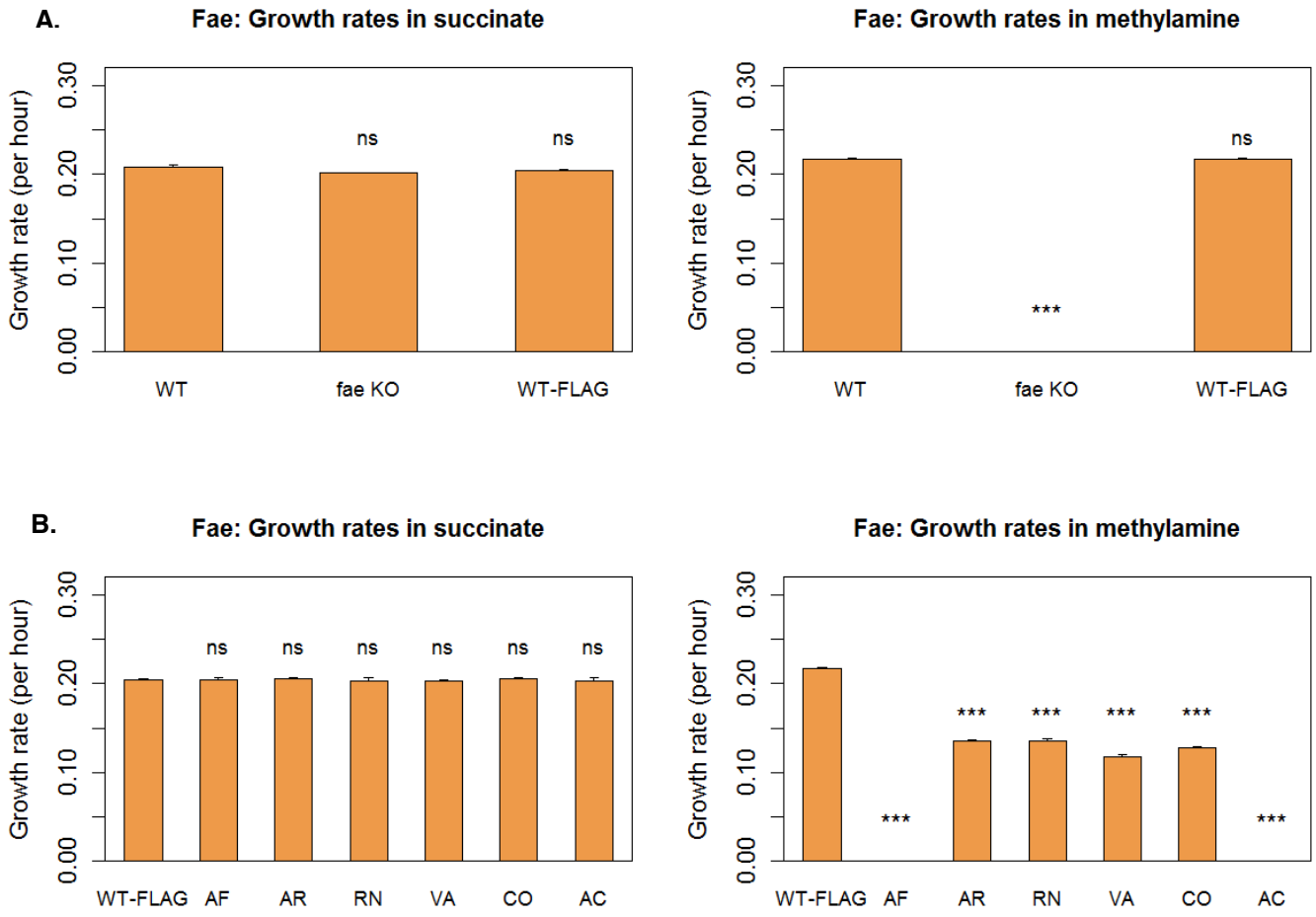

Figure 9: Growth rates of codon altered variants of *fae* in succinate and methylamine: (A) C terminal FLAG is neutral in *fae* as growth rate of WT-FLAG is not different from WT. The *fae* knockout does not grow in methylamine. Thus *fae* is an essential gene in presence of methylamine. (B) Growth rates of codon usage variants of *fae* do not differ from WT-FLAG in succinate but are significantly lower in methylamine. This indicates that changing codon usage of *fae* decreases fitness when the sole carbon source available in the environment is methylamine. Pairwise comparisons with WT (A) and WT-FLAG (B) using student's 2-tailed t test. *** p< 0.001, **p< 0.01, *p < 0.05, ns – not significant. n = 5 for all strains and error bars represent standard error of mean.

Growth rates of codon usage variants of *fae* (generated as a part of Agashe et al., (2013)) were measured in succinate and methylamine. The *fae* knockout (Δ*fae)* does not grow when the sole carbon source available is methylamine (Figure 9A) indicating that *fae* is an essential gene in these conditions. The WT version of *fae* with FLAG at C terminal grows just as well as the WT in both succinate and methylamine, indicating that the C terminal FLAG is neutral (Figure 9A). All the other codon-altered strains of *fae* do not show a growth defect in succinate but do grow slower than WT-FLAG in methylamine (Figure 9B. For detailed discussion see Agasshe et. al (2013)).

II. Generating strains with mauA alleles on the inducible plasmid pLC291

We have cloned AR, VA and CO into the inducible vector pLC291. We have set up triparental conjugations to transform this vector into Δ*mauA M. extorquens* AM1. Cloning the remaining alleles, WT, AF, RN and AC has failed multiple times and we are attempting to troubleshoot this. Once we have a full set *mauA* alleles on pLC291 we plan to measure growth rates to check if any growth defects show up.

III. Analysing background mutations in evolved strains with altered codon usage of fae

We analysed the genome sequences of the evolved *fae* lines using VarScan. Mutations called by the program in genes other than *fae* are summarized in Table 7 (SNPs unique to certain evolved *fae* lines), Table 8 (Indels unique to certain evolved *fae* lines) and Table 9 (SNPs and indels found in all evolved lines). The mutations in the *fae* gene itself were analysed by Agashe et al. (2016). We expected to find mutations in tRNA genes, translation related genes or other genes involved in the formaldehyde pathway of *M. extorquens* AM1. However, most of the mutations called by VarScan are in hypothetical proteins. Only one of the SNPs is near a tRNA gene. Hence it is likely that these mutations affect fitness in methylamine via epistatic interactions. These mutations are currently being confirmed via Sanger sequencing in the isolates that were sequenced as well as other isolates from the evolved lines. This will tell us whether the mutation has been fixed in the population or not. It is likely that mutations common to all evolved lines were already present in the ancestral strain used by Agashe et al. (2013) to generate codon usage altered *fae* strains and hence were propagated in the evolved lines. We will check if these common mutations are present in the *M. extorquens* AM1 Δ*fae* strain used by Agashe et al. (2013) to generate the codon usage variants.

Strain sequenced	Reference number/annotation	Genomic Position	Original Base	Mutated base	Frequency of SNP	Function	Sequencing results
eAR 11							
1	ACS39042.1	1225031	C	G	100	Conserved hypothetical protein; putative exported protein	
2	ACS40189.1	2480298	G	C	22.58	Hypothetical protein	Absent in eAR 11
3	pbpC	5390190	G	A	100	Penicillin-binding protein	Confirmed in eAR 11
eAF 11							
1	ACS39266.1	1463135	A	G	100	Conserved hypothetical protein	Confirmed in eAF 11
2	IGR	1839531	C	G	20.51	between etrahydromethanopterin-linked formaldehyde oxidation protein and formaldehyde-activating enzyme	
eAF 21							
1	ACS41615.1	4052202	C	G	37.04	Hypothetical protein	Absent in eAF 11
2	ACS41615.1	4052204	T	A	96.88	Hypothetical protein	Absent in eAF 11
3	IGR	3542311	G	A	99.88	between Conserved hypothetical protein (DUF88) and orotate phosphoribosyltransferase (OPRT) (OPRTase)	Absent in eAF 11
eAF 61							
1	ACS40444.1	2803840	T	C	100	Hypothetical protein	
eVA 11							
1	IGR	1839594	C	T	100	between etrahydromethanopterin-linked formaldehyde oxidation protein and formaldehyde-activating enzyme	Absent in eVA 11 and eVA 41
eVA 21							
1	ACS42902.1	5464128	A	G	99.59	conserved hypothetical protein; putative exported protein	Absent in eVA 21, eVA 41 and eVA 51 Present in eVA 11
eCO 11							
1	ACS43611.1	728644	G	A	99.74	Hypothetical protein	Absent in eCO 41
eAC 21							
1	IGR	1301033	G	T	62.06	btw hypothetical proteins	Absent in eAC 11
2	ACS38251.1	310646	T	G	31.58	conserved hypothetical protein	
3	arsR2	310675	C	G	26.41	transcriptional regulator, ArsR family, ArsR2	
4	arsR2	310677	C	T	23.5		
5	arsR2	310678	C	T	23.61		
6	arsR2	310690	G	C	28.78		
7	arsR2	310699	A	G	28.11		
8	arsR2	310705	C	T	25.43		
eAC 41							
1	ACS38251.1	310646	T	G	100	conserved hypothetical protein	
2	arsR2	310675	C	G	100	transcriptional regulator, ArsR family, ArsR2	
3	arsR2	310677	C	T	100		
4	arsR2	310678	C	T	100		
5	arsR2	310690	G	C	100		
6	arsR2	310699	A	G	100		
7	arsR2	310705	C	T	100		

Table 7: Summary of unique SNPs in evolved *fae* strains. eAF 11 means isolate 1 of replicate population 1 of the evolved strain AF and so on for all strains. IGR stands for intergeneic region. The mutations were called only if they were represented on both strands of DNA and the frequency of reads carrying the mutation was greater than 20%.

Strain sequenced	Reference number/annotation	Genomic Position	Original Base	Insertion/deletion	Frequency of indel	Function	Sequencing Result
eAR 11							
1	ACS39266.1	1463123	G	-CGCGGCCGACGA	79.94	Conserved hypothetical protein	Confirmed in eAR 11
2	IGR	1839530	G	-C	91.95	between fae and etrahydromethanopterin-linked formaldehyde oxidation protein	
eAR 21							
1	ACS41615.1	4052198	A	-GGTCGT	88.13	Hypothetical protein	Absent in eAR 11
eCO 11							
1	IGR	2938498	C	+AGCGGCTT	62.38	between tRNA gene and hypothetical protein	Absent in eCO 41
eAC 41							
1	ACS43260.1	348256	G	+GGCCGA	45.67	Hypothetical protein	

Table 8: Summary of unique Indels in the evolved *fae* strains. eAR 11 means isolate 1 of population 1 of evolved strain AR and so on for all strains. The '-' sign indicates a deletion and the '+' indicates an insertion in the Insertion/Deletion column. IGR stands for intergenic region. The mutations were called only if they were represented on both strands of DNA and the frequency of reads carrying the mutation was greater than 20%.

Discussion

There is no doubt that synonymous mutations can affect fitness though their effects are often weakly deleterious. All codon usage variants tested in this work show fitness defects. The *mauA* strains AR and RN have lower growth rates than WT-FLAG. This trend is similar to the one seen in *fae.* But we need to compare full sets of strains to confirm this. Contrary to expectations, VA and AC of *mtdA* show a growth defect only in succinate and not methylamine. AF also has a higher growth rate in methylamine compared to succinate. This indicates that altering codon usage of *mtdA* affects its function in succinate but not methylamine. It is possible that this could be the effect of the FLAG present at the N terminal of *mtdA.* We are currently attempting to generate WT-FLAG and AR strains of *mtdA* but have been unsuccessful in cloning these alleles into pCM433. Triparental conjugations for the remaining *mtdA* and *mauA* strains are underway. We also need to establish the neutrality of the FLAG in *mtdA.* From these preliminary results, it seems that the fitness effects seen in *mtdA* are dependent on the carbon substrate available. It would be interesting to see of this trend disappears when either the C terminal or the N terminal FLAG is removed. Once we have a full set of strains of each gene we can draw conclusions about whether the effect of synonymous mutations is context dependent.

Reference number/annotation	Genomic Position	Original Base	SNP/ Indel	Function	Sequencing Result
ACS38164.1	217318	C	T	putative mutator MutT protein precursor; putative NTP pyrophosphohydrolase	
ACS39608.1	1839422	T	C	etrahydromethanopterin-linked formaldehyde oxidation protein	Present in eVA 41 and VA 11
IGR	1839725	G	A	btw etrahydromethanopterin-linked formaldehyde oxidation protein and fae	Present in eVA 41 and VA 11
IGR	1839726	A	T	between two hypothetical proteins	Present in eVA 41 and VA 11
IGR	3159071	G	C	between two-component sigma-54 specific nitrogen transcriptional regulator, Fis subfamily; N-terminal response receiverand and putative helicase domain protein with DEAD/DEAH motif	
ACS40418.1	2777457	T	G	Hypothetical protein	
ACS40444.1	2803789	C	T	Hypothetical protein	
ACS41577.1	4001531	C	G	putative glutathione transferase (N-terminal fragment)	
ACS41577.1	4001531	C	G	putative glutathione transferase (N-terminal fragment)	
IGR	4160947	T	C	between two hypothetical proteins	
Pseudogene	482893	G	+C	Pseudogene	
IGR	2329710	A	-G	between hisI and a pseudogene	
ACS40660.1	3037768	T	-C	Hypothetical protein	
IGR	3159070	C	-G	between glng and putative helicase domain protein with DEAD/DEAH motif	
ACS41577.1	4001526	G	-CGTGC	putative glutathione transferase (N-terminal fragment)	
ACS43466.1	580985	G	+C	Hypothetical protein	
ACS43654.1	770862	T	-G	Superfamily I DNA and RNA helicase%2C helicase subunits-like protein (N-terminal part)	

Table 9: Summary of SNPs and Indels common across all the evolved *fae* strains that were sequenced. The '-' sign indicates a deletion and the '+' indicates an insertion in the SNP/Indel column. No sign indicates the mutation is a SNP. IGR stands for intergenic region. The mutations were called only if they were represented on both strands of DNA and the frequency of reads carrying the mutation was greater than 20%.

Results from Agashe et al. (2013), Hauber et al. (2016) and our preliminary analysis of whole genome sequence data of the evolved *fae* lines indicates that tRNA-codon usage imbalance may not explain the fitness effects seen in our experiments. This may be due to the fact that we only alter the codon bias of a gene and not the entire genome.

Napolitano et al. (2016) attempted to replace all occurrences of two rare arginine codons with alternative synonymous codons in essential genes of *Escherichia coli* using an automated recombineering approach. They failed to make certain

synonymous mutations that were either near the start codon or the stop codon of essential genes. To work their way around this, they instead made substitutions that conserved the mRNA secondary structure or the similarity to the SD sequence. Their strain also accumulated additional mutations. One of these was in the tRNA used to decode the codon they were replacing the rare codons with ($tRNA^{Arg}$ *argU*). However, when this mutant tRNA was replaced with the wild-type tRNA, they did not observe any effect on fitness. They found that mRNA secondary structure and affinity to Shine-Dalgarno like sequences are correlated with successful generation of mutants. Another study attempted to replace 13 rare codons occurring in ribosomal genes of *E. coli* with other synonymous codons (Lajoie et al., 2013). Ribosomal genes were selected as they are highly expressed and are expected to be codon optimized. They were successful in replacing these 13 codons in certain genes whereas this failed in others. This too supports the hypothesis that the effect of synonymous mutations is context dependent. Also, replacing rare codons of ribosomal genes with other synonymous codons caused a decrease in growth rate.

Synonymous mutations occurring near the 5' end of essential genes have larger effects on fitness. Goodman et al. (2013) made a large number of synonymous variants with varying codon usage at the N terminal of super folding GFP and quantified gene expression. A weak correlation with SD like sequences and a strong correlation with mRNA folding energy was observed. Agashe et al. (2016) evolved the *fae* codon usage variants till they recovered fitness to almost wild-type levels. They identified SNPs arising around the 5' end of *fae* that compensated the fitness defects of their codon usage variants.

From literature as well as our preliminary results, it does seem like the effect of synonymous mutations is context dependent due to multiple mechanisms mediating how they affect fitness. Though it seems difficult to disentangle the causes of this effect, we hope to achieve this by comparing fitness data of codon usage altered genes within the same pathway.

References

Agashe, D., Martinez-gomez, N.C., Drummond, D.A., and Marx, C.J. (2013). Good

Codons , Bad Transcript : Large Reductions in Gene Expression and Fitness Arising from Synonymous Mutations in a Key Enzyme. Mol. Biol. Evol. *30*, 549–560.

Agashe, D., Sane, M., Phalnikar, K., Diwan, G.D., Habibullah, A., Martinez-Gomez, N.C., Sahasrabuddhe, V., Polachek, W., Wang, J., Chubiz, L.M., et al. (2016). Large-Effect Beneficial Synonymous Mutations Mediate Rapid and Parallel Adaptation in a Bacterium. Mol. Biol. Evol. *33*, 1542–1553.

Bailey, S.F., Hinz, A., and Kassen, R. (2014). Adaptive synonymous mutations in an experimentally evolved Pseudomonas fluorescens population. Nat. Commun. *5*, 4076.

Bosch, G., Skovran, E., Xia, Q., Wang, T., Taub, F., Miller, J.A., Lidstrom, M.E., and Hackett, M. (2008). Comprehensive proteomics of Methylobacterium extorquens AM1 metabolism under single carbon and nonmethylotrophic conditions. Proteomics *8*, 3494–3505.

Bulmer, M. (1988). Are codon usage patterns in unicellular organisms determined by selection-mutation balance? J. Evol. Biol. *1*, 15–26.

Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. Genetics *129*, 897–907.

Chubiz, L.M., Purswani, J., Carroll, S.M., and Marx, C.J. (2013). A novel pair of inducible expression vectors for use in Methylobacterium extorquens. BMC Res. Notes *6*, 183.

Curran, J.F., and Yarus, M. (1989). Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. J. Mol. Biol. *209*, 65–77.

Delaney, N.F., Kaczmarek, M.E., Ward, L.M., Swanson, P.K., Lee, M.C., and Marx, C.J. (2013). Development of an Optimized Medium, Strain and High-Throughput Culturing Methods for Methylobacterium extorquens. PLoS One *8*.

Goodman, D.B., Church, G.M., and Kosuri, S. (2013). Causes and Effects of N-terminal Codon bias in Bacterial Genes. Science (80-.). *342*, 475–480.

Gouy, M., and Gautier, C. (1982). Codon usage in bacteria: Correlation with gene

expressivity. Nucleic Acids Res. *10*, 7055–7074.

Hauber, D.J., Grogan, D.W., and DeBry, R.W. (2016). Mutations to less-preferred synonymous codons in a highly expressed gene of Escherichia coli: Fitness and epistatic interactions. PLoS One *11*, 1–16.

Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. *2*, 13.

Kelsic, E.D., Chung, H., Cohen, N., Park, J., Wang, H.H., and Kishony, R. (2016). RNA Structural Determinants of Optimal Codons Revealed by MAGE-Seq. Cell Syst. *3*, 563–571.e6.

Kimura, M. (1968). Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. *11*, 247–269.

King, J.L., and Jukes, T.H. (1969). Non-Darwinian evolution. Science (80-.). *164*, 788–798.

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., Mclellan, M.D., Lin, L., Miller, C. a, Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2 : Somatic mutation and copy number alteration discovery in cancer by exome sequencing VarScan 2 : Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. *22*, 568–576.

Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-Sequence Determinants of Gene Expression in Escherichia coli. Science (80-.). *324*, 255–258.

Lajoie, M.J., Kosuri, S., Mosberg, J.A., Gregg, C.J., Zhang, D., and Church, G.M. (2013). Probing the Limits of Genetic Recoding in Essential Genes. Science (80-.). *342*, 361–364.

Li, G.-W., Oh, E., and Weissman, J.S. (2012). The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature *484*, 538–541.

Lind, P.A., and Andersson, D.I. (2013). Fitness Costs of Synonymous Mutations in the rpsT Gene Can Be Compensated by Restoring mRNA Base Pairing. PLoS One

8, 1–6.

Lind, P.A., Berg, O.G., and Andersson, D.I. (2010). Mutational Robustness of Ribosomal Protein Genes. Science (80-.). *330*, 825–827.

Marx, C.J. (2008). Development of a broad-host-range sacB-based vector for unmarked allelic exchange. BMC Res Notes *1*, 1.

Marx, C.J., and Lidstrom, M.E. (2004). Development of an insertional expression vector system for Methylobacterium extorquens AM1 and generation of null mutants lacking mtdA and/or fch. Microbiology *150*, 9–19.

Marx, C.J., Chistoserdova, L., and Lidstrom, M.E. (2003). Formaldehyde-Detoxifying Role of the Tetrahydromethanopterin-Linked Pathway in Methylobacterium extorquens AM1. J. Bacteriol. *185*, 7160–7168.

Napolitano, M.G., Landon, M., Gregg, C.J., Lajoie, M.J., Govindarajan, L., Mosberg, J.A., Kuznetsov, G., Goodman, D.B., Vargas-Rodriguez, O., Isaacs, F.J., et al. (2016). Emergent rules for codon choice elucidated by editing rare arginine codons in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. *113*, E5588-97.

Plotkin, J.B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. TL - 12. Nat. Rev. Genet. *12 VN-r*, 32–42.

Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al. (2015). Codon optimality is a major determinant of mRNA stability. Cell *160*, 1111–1124.